

## **General Disclaimer**

### **One or more of the Following Statements may affect this Document**

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

"Made available under NASA sponsorship  
in the interest of early and wide dis-  
semination of Earth Resources Survey  
Program information and without liability  
for any use made thereof."

NSG-5080  
7.6-10439  
CR-148516

APPLICATION OF LANDSAT DATA TO AGRICULTURAL  
RESOURCE PROBLEMS WITH EMPHASIS ON THE  
NORTH AMERICAN GREAT PLAINS

Progress Report

February 28, 1976

(E76-10439) APPLICATION OF LANDSAT DATA TO  
AGRICULTURAL RESOURCE PROBLEMS WITH EMPHASIS  
ON THE NORTH AMERICAN GREAT PLAINS Progress  
Report (Columbia Univ.) 34 p HC \$4.00

N76-29669

Unclas  
00439

CSCL 02C G3/43

Principal Investigators:

Kempton Webb  
Colin J. High *etc*  
Jerry C. Coiner

## CONTENTS

### **I. Content of the Research**

### **II. Experiments in Progress**

#### **A. Support to GISS 24-Channel and Thematic Mapper Studies**

#### **B. Environmental factors affecting the utility of LANDSAT data for crop classification**

##### **1. Soil color variation**

##### **2. LANDSAT analysis of small field agriculture**

#### **C. Environmental factors affecting crop productivity**

##### **1. Saline seeps in North Dakota**

##### **2. Hail damage to crops**

#### **D. Landscape analysis (Tropical land cover changes and human activity)**

##### **1. West African Sahel-Savanna**

##### **2. Upper Amazon Settlement**

##### **3. Tropical land component analysis**

### **III. Planned Research**

## I. Content of the Research

As a method of data acquisition for agricultural and environmental monitoring, remote sensing has one feature which is simultaneously its greatest strength and most intractable weakness. Data acquisition is essentially holistic. Sensors provide only one estimate of reflected or emitted energy from a fixed areal unit on the surface and, to varying degrees, the atmospheric cone interposed between the sensor and the target. Since all landscape components in the areal unit contribute to the energy reflected or emitted from that spatial field, the energy return at the sensors is a complex expression of all landscape components, natural and manmade, occupying the field. In the past, analysis of remotely sensed data attempted to take this single or multiple radiometric attribute of the landscape unit and use it to characterize (describe) one of the components in the landscape. Such attempts met with varying degrees of success, predicated on two classes of analytically controlled factors: 1) the degree to which the landscape component dominated the spectral or radiometric character of the return, and 2) the goodness of fit of matter/energy interactions with the typology containing the landscape component.

In most remote sensing research, these two factors received little, if any, consideration. Existing typologies were used without reference to relationships between landscape

components and matter/energy interactions. For instance, it seems futile to attempt floristic classification of vegetation by using remote sensing data because the Linnaean classification is based on the morphological similarities between the flowering parts of the plant, parts which normally make up a small portion of the total plant biomass that is sensed. Typology problems are compounded by the way in which landscape components interact to create the return. Most research to date has failed to account for the results of complex interactions and the spatial variations which accompany the changing nature of these interactions.

The Columbia University, Department of Geography, is attempting to investigate different human-induced and natural environmental phenomena using remotely sensed data, but with full appreciation for the above problems. Currently being researched are aspects of data analysis encompassing a single unique descriptor for an environmental variable and the development of new landscape typologies, specifically designed to incorporate matter/energy interactions. The program is now underway to extract predefined man-made and environmental variables from LANDSAT and other sensor data. In addition, effort is being directed to the formulation of new land cover typologies using the concepts of numerical taxonomy.

## II. Experiments in Progress

### A. Support to GISS 24-Channel/Thematic Mapper Studies

Since October, 1975, when initial sample of 24-Channel multispectral scanner (MSS) data first became available to GISS, Columbia researchers under grant NSG 5080 have been involved in the analysis of this data. Columbia researchers have participated in correlation studies and the initial unsupervised and supervised classifications of the MSS data from the LACIE intensive test site 1960 in Finney County, Kansas. Field observation of the site in early November indicated that in some cases, misclassification of crops from the 24-Channel data resulted from mistaken identification in the original survey or from unrecorded crop condition variations within the fields.

Since December, 1975, the Geography Department's resources have been applied mainly to the problems of developing, acquiring and implementing ground truth files for comparison with 24-Channel data from the LACIE Finney and Williams County test sites.

Through the end of February, full resolution ground truth files have been created for two flight lines, one over the Finney County test site (FL 1, 6/9/75) and one from the Williams County test site (FL 2, 6/22/75). These two flight lines of ground truth data (each consisting of over one million pixels) will provide GISS with data to support classification

of 24-Channel data at full resolution and develop figures of merit for the classifications. These classifications will aid in analysis of the impact of changing sensor system's parameters, such as IFOV (resolution) and band configuration on the ability of the system to provide effective data on crops and crop acreage.

In addition to the two full resolution ground truth arrays, the Columbia research group has developed six ground truth files for LARS' spatially degraded 24-Channel data at 30 meter and 60 meter resolutions. These arrays require the coding of over 150,000 discrete ground cells into computer readable format, punching, verification and loading as computer files. These tasks were undertaken as part of the work necessary to provide an analytical data base to support classification of spatially degraded 24-Channel data developed to simulate the Thematic Mapper instrument's output. The spatially degraded data are intended to simulate how classification accuracy varies as a function of changing the spatial resolution from 60 meters to 30 meters. Research personnel funded under this grant are also taking part in the classification phases of the project in addition to the preparation and management of ground truth files.

B. Environmental factors affecting the utility of  
LANDSAT data for crop classification

1. Soil color variation

LACIE has identified a problem in the derivation of crop signatures from LANDSAT, in that crop signatures are spatially

variant causing signature reliability decay. This variation tends to limit a given crop's unique signature to a small area and to require extensive ground data for reformulation of that crop's signature to make it applicable in other areas.

To investigate the role of soil color changes in this signature reliability decay, the LACIE site at Saline County, Kansas (Segment No. 1962) was chosen for soil color analysis. The Soil Conservation Service 1:31,680 scale soils map and accompanying soil survey were used in conjunction with LACIE ground truth to create three virtual pixel maps of training sets. These maps are crop type, Munsell soil color dry, and Munsell soil color moist. Preliminary computer analysis using signature derivation demonstrated that direct crop type classification is ambiguous, with signatures for given winter wheat fields showing large variations in radiometric values in all four bands (standard deviation greater than 20%). When signatures were developed for soil color types, the variation within class was of the same order (standard deviation greater than 20%), suggesting a relationship may exist. Continued experiments will structure a combined soil color/crop type virtual map for the test site, thereby using soil color to define strata within each crop type or field condition. These experiments will lead to an understanding of the role soil color plays in crop signatures and will define the sensitivity of crop signatures to changes in soil colors.



It is expected that the experiments will also provide data on whether it is necessary to account for rainfall-induced soil color changes. If classification of crops are significantly improved for the Saline site, the experiment will be extended to the remaining four LACIE test sites in Kansas. Some assessment of other soil parameters, such as texture and structure, will then be undertaken. Future research would envision the development of an environmental factors matrix of soil parameters, rainfall occurrence and micro-relief, which could be accessed to determine winter wheat signature variation by environmental conditions.

## 2. LANDSAT analysis of small field agriculture

An area in Southern New Jersey was evaluated to assess the feasibility of a test site within easy reach of the project headquarters in New York City. The purpose of this site is to make possible detailed checking of the classification algorithm for LANDSAT data in an area where a wide range of agricultural practices can be kept under close observation during all seasons of the year at relatively low field work cost.

The area is approximately fifty square miles centered on the village of Seabrook in Cumberland/Salem Counties, Southern New Jersey (see attached location map). Topographic maps, soil maps, air photographs as well as LANDSAT and SKYLAB coverage are available for the site. The area is

gently rolling agricultural land located predominantly on the interfluves with patches of woodland in the valleys.

A preliminary crop inventory in late September identified the following crops: corn, soybeans, winter wheat, alfalfa, hay, and vegetables. These crops are mostly grown in medium-sized fields of about 20-40 acres, although there are many smaller fields. The crop combination is similar to that of the Midwest and should provide a suitable test site to critically evaluate the effectiveness of the classification procedures which are being used in other areas. The occurrence in part of the test site of small field sizes and intensive vegetable growing (truck farming) will allow testing of methods for crop classification and inventory from LANDSAT of small field agriculture similar in scale to that found in the subsistence agricultural systems in developing countries.

To date, work on this project has included a preliminary qualitative field inventory of crop types, accumulation of background data maps and air photographs, examination of LANDSAT imagery and initial digital processing of gray scale printouts of four band MSS LANDSAT data from the test site. Detailed ground truth collection by field crop mapping and computer classification of available LANDSAT data will follow. It is proposed to continue the work on this test site for at least one full year matching ground truth collection with LANDSAT passes as closely as possible.

## C. Environmental factors affecting crop productivity

### 1. Saline seeps in North Dakota

In conjunction with the Soils Department, North Dakota State University, studies have been undertaken to use LANDSAT and GISS aircraft spectrometer data in the analysis of saline seeps. Saline seeps are an artifact of glaciation. When the glacial till was laid down, lenses of saline sands were imbedded into the glacial substrate on which recent soils have developed. When the native vegetation (short grass prairie) is replaced by messicol vegetation (stripped crop spring wheat), water that would normally be transpired by the native grasses percolates into the lenses of saline material and is transported laterally, seeping to the surface in low lying areas within the rolling plain. Whenever water that has been transported through these lenses reaches the surface of the land, both native and messicol vegetation is killed because of the high salinity of the water. The vegetation is then replaced by halophytic forms. Essentially these seeps are a human-induced geomorphic process caused by strip cropping, in that unvegetated areas in fallow allow water to reach the saline lenses. Understanding the causes of saline seeps and the extent and distribution of the affected area should allow North Dakota State University to make better recommendations to the farmer concerning practices which might control the problem. Recommendations such as how much area to leave fallow at spring planting depend on knowing the dynamics of the seeps and how they expand and contract with annual moisture changes.

To support research on the nature, extent, and spread of saline seeps, a series of experiments are now being undertaken. LANDSAT data are being studied for a twenty-four section area in Stark County, North Dakota. Ground truth has been collected for signature development and a larger area has been delimited for survey to test signature reliability. This research is intended to identify the time of year the seeps are most detectable and to determine how the area of seep influence detected in the data corresponds to the ground survey measurements. If this phase is successful, the signatures will be applied to all of Stark County (1,000 square miles) to determine if the area affected by the seeps is expanding or contracting from 1972 to the present. The affected areas will also be mapped from LANDSAT to aid field workers in their location.

Data from the GISS airborne spectrometer is also being used to study the seeps. On October 11, 1975, three instrumented saline seep areas in Divide and Williams County, North Dakota were flown with the spectrometer. These missions were supported by extensive ground data collection by North Dakota State University. The purpose of these flights was to use spectral data on the seeps, lenses, halophytic vegetation and uplands to aid in understanding the ground water/geomorphological complex which creates them. If subtle spectral variations in ground cover or soil color can be found which define the upland water source areas above the seeps, it may

be possible to use spectrographic data to predict the occurrence of the seeps before their break out or at a very early stage. This predictive capability would allow conservation measures by farmers to take place before crop acreage is lost to production.

The October ground and spectrometer data are now being processed and creation of files for spectral analysis of the seeps is underway. Preliminary experiments will concentrate on seep definition within the spectrometer data. If this is successful, efforts will be made to determine unique spectral characteristics which delimit the path of movement from the upland water source to the point where the seep surfaces.

## 2. Hail damage to crops

In the North American Great Plains, more than 235 million dollars worth of field crops (mostly cereal grains) are lost to hail yearly. Yet, little is known about the pattern of hail damage over large areas. Also, measurement of the extent of damage is accomplished only for those areas where crop insurance is in effect (both Federal and private insurance is available). To better understand the spatial distribution of hail damage from severe storms and to assess the damage the storms inflict on crops and natural vegetation, a series of LANDSAT experiments will be undertaken.

At present, a survey is being conducted to determine the state of the art in hail damage assessment. Experiments

will determine if hail storm patterns are detectable in LANDSAT data, both by image interpretation and digital processing methods. Efforts will then be directed to the assessment of the severity of damage.

D. Landscape analysis (Tropical land cover changes and human activity)

1. West African Sahel Savanna

In developing countries predominantly located in tropical regions, the pace of environmental change has increased markedly as the result of rapid population growth and the introduction of new technologies. It is in these areas that agricultural developments are most critical if the critical food needs of the developing world are to be met. Existing methods for collection of agricultural and environmental data have not in general kept pace with the rate of change. LANDSAT is in many respects ideally suited for the monitoring of agricultural and environmental change on a regional or sub-continental scale. In many areas of the developing world, it is the most up-to-date source of data available.

The purpose of the West African Sahel-Savanna project is to develop methods for rapid automated processing of LANDSAT data for agricultural and environmental monitoring in the dry Sahel-Savanna regions of Sub-Saharan West Africa. Although research is concentrated in this area, it is anticipated that the methodology will be applicable to other areas of the tropics. The small sized fields, coupled with complex

intercropping, fallowing and rotation, make it virtually impossible to apply the methods of analysis of LANDSAT data designed for the large field, monospecific cropping regions of North America. It is proposed, therefore, to concentrate upon using LANDSAT data to map land cover assemblages which represent particular agricultural land uses and to develop models based on field work and detailed air photo interpretation which can be used to predict crop production within each land cover assemblage.

Preliminary site selection work has begun on three sites in the dry Savanna areas of North Nigeria. Northern Nigeria was chosen as representative of the dry Savanna in West Africa which is undergoing relatively rapid agricultural change. Background data, maps and air photographs are available, and the investigators have contacts with Nigerian scientists working in the area.

Three sites shown on the accompanying map have been selected.

Soba Site: Preliminary analysis of LANDSAT data by digital processing has been completed for a site centered on the village of Soba (population 4,000) on the Zaria-Jos Road in North Central State. Classification of land cover by digital processing has been compared with a land use map produced by field survey and air photo interpretation in 1973-1974. Agreement between the LANDSAT data classification and land use map has been relatively poor but careful investigation

of the field-produced land use map indicates that this poor agreement between it and the LANDSAT-derived classification can be attributed, at least in part, to the unsuitability of the land use categories for this purpose. There is clearly a need to develop appropriate land cover classifications for use in conjunction with remote sensed data in tropical areas.

Sokoto Valley Site: Initial analysis has been made by both image interpretation and digital processing of LANDSAT data for a site of the West Bank of Sokoto Valley near Argungu in the North West State of Nigeria. Characteristic land cover patterns of forest clearance cropping around nucleated villages have been recognized and compared with 1:50,000 maps of the area. Supervised training and survey by digital methods using the GISS algorithm is in progress. A preliminary report is appended as Attachment 1.

Katsina Site: A trial site has been selected for detailed study north of Katsina in North Central State of Northern Nigeria. Although no detailed work has been conducted as yet, a preliminary interpretation of the available LANDSAT images for the area indicates that it has an ideal combination of land cover types including agriculture, settlements, dry forest, grassland, and valley bottom lands (Fadamas) to provide a good test site for developing machine processing methods for this environment. In addition, good quality recent air photographs on a scale 1:40,000 have been taken in the dry season (December 1974). These should be suitable for



compiling a base map for ground truth collection. The area is accessible for field work on all-weather roads.

Contacts have been made with Nigerian scientists and an agreement to provide local assistance has been reached. The Nigerian Federal Survey Department has been asked to authorize the release of air photographs necessary for field work on the Katsina site. As soon as these air photographs are available, it is proposed to begin interpretation of Katsina site to collect supporting data for digital processing. This will proceed in parallel with further processing of LANDSAT data as new coverage becomes available. Work will continue on developing an appropriate land cover classification method for tropical subsistence agriculture to use in conjunction with remote sensing.

## 2. Upper Amazon settlement

In cooperation with Eric and Jane Ross, Columbia University, Department of Anthropology, an analysis of LANDSAT data to locate potential settlement sites of the Achuar Indians on the Upper Amazon River has been undertaken. To date, a number of potential village sites have been identified using ground data collected by the Rosses and a team of Italian anthropologists. These data were used to develop a signature thought unique to the village environs. Using this signature, several scenes of LANDSAT data acquired 29 September 1973 were then classified, revealing several potential village sites which met predefined environmental criteria established

by Jane and Eric Ross. Some of the LANDSAT-identified sites have been discussed with the Italian anthropologists, now en route to the area, and these anthropologists will attempt to check the reliability of the locations. Other locations will be checked by the Rosses during field studies in the summer of 1976. Further research will attempt to define ecological zones within the settlement areas, for anthropologists have found that primitive settlement locations are ecologically controlled, i.e., humans settle where food and fiber are most accessible. In many cases, this has meant that settlements tend to be at seams or boundaries in the landscape, allowing the inhabitants to exploit the resources of two or more different ecological zones. Using LANDSAT data, studies will be undertaken to identify these ecological zones and their boundaries. Such studies will support not only settlement location efforts, but also human territoriality investigations and the interaction between primitive human settlements.

### 3. Tropical land component analysis

One of the last remaining resource pools on Earth is the tropical rainforest in the Amazon Basin. This area will come under heavy development pressure in the next few decades. If some effort is to be made to renew the Amazon resource base and to sustain its potential productivity, the components in their pristine state must be analyzed now before extensive development modifies them. In an initial attempt

to begin this analysis, LANDSAT data will be used to determine if predefined classes of vegetation and water can be assigned unique spectral classes, so that landscape changes can be monitored. These experiments would attempt further to define new landscape components which have not previously been identified because of the general inaccessibility of the area. In order to conduct the research, a system of analysis particularly suited to the sensor systems currently available and to the humid tropical rain forest environment will be developed.

At present, imagery and tapes on hand at GISS are being studied to define areas where previous ground studies have been conducted. These studies will provide data to support the development of signatures for the various vegetation and water types in the Amazon basin. Using these signatures, studies in vegetation/water type mapping as well as surveys for new landscape components will be undertaken.

### III. Planned Research

During the next six months period it is envisioned that the majority of the Columbia effort under the grant will remain focused on support of GISS 24-Channel/Thematic Mapper studies. This support will take the form of providing and managing ground truth arrays and conducting digital analysis to determine crop types and acreages.

In addition, this research project has identified the need to develop land cover typologies specifically suited for use with remote sensed data and which are appropriate for policy makers and resource managers. The pressing requirements for immediate application of remote sensing data can best be met through the development of new typologies rather than through continually "brute forcing" the data into existing typologies. Two approaches will be used to define new typologies which will enhance the use of remotely sensed data. (i) Identification and association of innate categories within the data with preference to specific planning and policy information requirements. (ii) The development of new land cover typologies which reflect the physical characteristics of the land cover. Both of these approaches will utilize the techniques of numerical taxonomy.

**ATTACHMENT 1**

**DIGITAL PROCESSING OF LANDSAT DATA  
FOR THE SURVEY OF RURAL LAND USE  
IN THE WEST AFRICAN SAVANNA**

**by**

**Helene Wilson and Colin J. High  
Department of Geography  
Columbia University, New York**

**Paper presented at the annual meeting of the  
Association of American Geographers  
New York City, April 1976**

## ABSTRACT

A case study involving the use of digital processing of LANDSAT data to identify rural land use in an area in Northwestern Nigeria points to the potential of the satellite as a key source of information on land use in the tropics, where for the most part such data is lacking. Three land classes -- savanna woodland, cultivated land, and villages -- were distinguished, although the distinction between woodland and villages could be made only with the addition of spatial information. Digital processing of LANDSAT data for the study of land use in the tropics offers numerous advantages over conventional image interpretation techniques. Among these advantages are maximum utilization of available satellite data both in terms of spectral information and spatial resolution, timeliness of results, and efficiency in dealing with large quantities of data. However, the necessarily objective and precise nature of the classification algorithms suggests the need for rethinking existing classification schemes for tropical land use.

## INTRODUCTION

The National Aeronautics and Space Administration LANDSAT program (formerly Earth Resources Technology Satellite, ERTS program) has provided geographers, among other scientists, with a valuable source of data of near global scope. Most work by geographers to date has involved interpretation of the LANDSAT data from the printed or transparency image either as single band images or in the form of color composite images. The techniques of image interpretation developed from air photograph studies have been employed in a modified form to data from LANDSAT images. However, by using automated digital data processing to analyze the original LANDSAT sensor data from the recorded magnetic tapes, a number of advantages can be gained. The major advantage of this method is that it utilizes all the information recorded and transmitted from the satellite sensors rather than just the degraded data of the image. Furthermore, the original data can be manipulated as desired using any combination of the four bands of the multispectral scanner that is most appropriate. At the same time the data may be examined at the scale of the ultimate resolving power of the sensor, namely one pixel of 60 by 80 meters. LANDSAT collects so much data that only a fraction of it could ever be analyzed by manual techniques. Therefore, if the full potential of the system is to be exploited, automated data processing will be essential.

Other advantages of a satellite-based remote sensing system -- global scope, consistency, repeated observations, and large area of coverage -- make LANDSAT a valuable complement to other remote sensing methods and ground survey in those areas such as North America and Western Europe, where these other data sources are well developed. In those areas of the world, especially in the developing

countries of the tropics where alternative sources of data are scarce and air photographs outdated or nonexistent, LANDSAT data is especially valuable. In some cases it may be the only up-to-date remote sensing data and in others the preferred data source for many applications. For studies of agriculture as well as of vegetation alteration, overgrazing, and desertification over large areas of the tropics, rapid survey by automated processing of LANDSAT data may be very valuable. The technology has the potential of providing information quickly enough so that it may be of value to national and international organizations seeking to deal with regional scale disasters such as the recent African Sahel drought.

Substantial problems remain to be solved before the state of the art reaches the level where it can respond to the needs generated by a major environmental crisis. These problems fall into three groups. First it is necessary to develop appropriate methods for computer analysis of the LANDSAT data on the land cover types of the tropics. Procedures must be developed appropriate to the particular environmental and agricultural conditions prevailing. This task is complicated by the relatively small field size, intercropping, and fallowing patterns of tropical subsistence agriculture. The monospecific large fields of North American agriculture are much easier to classify by contrast.

The second problem lies in the fact that the existing classification schemes for land use and vegetation in the tropics are by no means consistent. Thus we are faced with the necessity of translating classifications based on explicit quantitative criteria in our algorithms into field-derived classifications of land types based on much less exacting criteria. It is anticipated that a much more exact numerical approach to land classification may have to be adopted before the full potential of remote sensing data can be realized.



Finally, as is evident from the study described in this paper, the LANDSAT coverage is frequently inadequate for full analysis. The lack of coverage is most acute for the rainy season in the tropics. For many parts of West Africa, for example, there are no good quality wet season satellite data available. The problem is, of course, partly because of the high incidence of cloud cover in the wet season but it is also a function of the data recording system on the satellite. In areas within range of a receiving station the data are transmitted directly to the ground as acquired. Beyond the receiving station range the data are recorded on tape on board the satellite and relayed to the ground when next in range. The on-board recording capacity is insufficient to record all the out-of-range data worldwide which the sensors could acquire, so the system must be selective. It is generally programmed to omit areas during expected periods of extensive cloud cover. This situation will persist until additional receiving station is just being completed at Kinshasa in Zaire by ERTS-Zaire and this will be able to receive all LANDSAT data transmissions while the satellite is over tropical Africa. Data availability for Africa should therefore improve considerably.

#### METHODOLOGY

The procedure used in this study to extract the desired information from LANDSAT data stored on computer tapes is based on an algorithm developed at the Goddard Institute for Space Studies (GISS). First, one must locate the area of interest on a black and white or color composite LANDSAT image (scale: approximately 1:1,000,000) in order to define the area for which a computer grey scale printout is to be generated at a nominal scale of 1:21,000. The grey

scale is used for more precise delimitation of both the training site and the test area. The training site is used to determine the signature of the land category being investigated. This signature is then tested for its validity by its extension to another area.

If detailed ground truth is available, signature training may be accomplished by creating a ground truth array that is registered with the computerized LANDSAT data. The signature of the desired category is then, for each band, the average of the radiance values for all picture elements (pixels) which fall within the desired ground category. If information about the area of investigation is scanty, as was the case in this study, an alternative method may be used to develop the training signature: unsupervised classification may allow one to separate out the pixels in that ground type and compute its signature as described above. The unsupervised classification, however, is useful only if the approximate location and/or areal extent of the ground category is known (e.g., from maps) and if the ground category is sufficiently differentiated in the LANDSAT data from its surroundings. The extent of this differentiation depends both on the data itself and on the nature of the unsupervised classification algorithm. The training signature, once determined, is then used as input into a supervised classification on the training site itself to determine the parameters needed to classify out the previously identified training pixels.

The supervised classification algorithm developed at GISS treats each pixel as a vector in a "color" space with four axes corresponding to the four bands of the LANDSAT multispectral scanner. The direction of the vector denotes the "color balance," whereas the length of the vector represents "albedo" or brightness -- that is, the sum of the radiances in all bands. Vectors representing

similar pixels (i.e., ground types) will tend to cluster in "cones" with different axes. The axis of any cone can be thought of as the representative signature for all pixels whose vectors fall within that cone.

The parameter which describes the angle from the origin that encloses the cone is one of the parameters which must be determined in the supervised training classification. Other parameters may be adjusted in order to refine the training signature. Ground categories that differ from each other more in brightness than in color balance may be more successfully extracted from the data by increasing the relative weighting of albedo. Bands may also be differentially weighted or even discarded altogether if it appears that the discriminating ability lies in fewer than four bands. In addition, hyperspaces with more than four axes may be used for multi-temporal classification. (See Fig. 1 for a two-space representation of the classification algorithm.)

The final step is the extension by supervised classification of the training signature with the appropriate parameters to the test area. Success is measured by the extent of the match between the selected cone and the real world distribution of vectors representing the desired ground category.

### CASE STUDY

The study area, approximately 50 by 50 kilometers, is situated in Northwest Nigeria in North Western State at approximately 13°N and 4°30' E (Fig. 2). This area lies in the Sudan Savanna vegetation belt in a region markedly affected by the recent drought but just to the south of the severely stricken Sahel zone. The natural vegetation is predominantly grasses and shrubs with scattered trees. The study area includes extensive "forest reserves" of dry savanna woodland broken by large areas

of permanently cultivated land around the villages with scattered shifting cultivation at further distances from the villages. Annual rainfall averages approximately 730mm (28in), with most of the precipitation occurring between May and September. The dominant ethnic group is the Hausa, whose settlement pattern consists primarily of nucleated compact villages, the composition of which is mud, thatched roofs, soil, trees, and small gardens.

LANDSAT data used in the study was from an overpass (no. 2014-09190) on February 5, 1975, in the middle of the dry season. Information about village location was obtained from the 1:100,000 Nigerian Federal Surveys topographical maps (sheets 1, 2, 8, and 9).

The study area contains 44 individual villages and clusters of villages identified from the maps. Two of the larger villages were selected for training. Despite the possibility of introducing bias due to differences in composition between large and small villages, it was decided that it was preferable to use a training signature which represented a relatively large group of village pixels.

Since the maps indicate only a point location for the villages, the unsupervised classification routine was used to identify the pixels included in each of the two villages. Although the two villages yielded very similar signatures, histograms revealed that the distribution of radiances in each of the four bands was unimodal in one village and bimodal in the other, possibly reflecting the presence of a waterhole in the latter village. Three training signatures were developed: a separate one for each village and a composite signature based on the two separate signatures.

The albedo-color weighting that yielded the best results in the supervised training was one that weighted albedo somewhat higher than is usual in comparable studies of rural land use in temperate zones. Previous work by the authors in Northern Nigeria suggests that albedo may be a significant discriminant in semi-arid regions, especially during the dry season.

Due to the paucity of detailed ground truth, the objective of signature extension was to test the ability of the training signature to identify village sites rather than to pick out every village pixel. That is, if the signature classified out only one pixel from what might be a five-pixel village, this would be considered successful. The three training signatures yielded similar results on testing. The locations of village settlements were clearly identified and villages were separated from surrounding cultivated land in many cases. However, the results suggest a strong similarity between the signatures of villages and of naturally vegetated savanna woodland areas. As a result of this spectral confusion, spatial criteria had to be applied to the classification results to complete the identification of village sites. A village located at least 1.5 kilometers from naturally vegetated areas and at least .5 kilometers from other villages could be most clearly identified with the available maps. Villages located in clusters and/or contiguous with naturally vegetated areas could not be clearly identified because of confusion with the background class of land cover. It is possible that some of the villages have been abandoned since being mapped or that a number of them are temporary settlements and hence significantly different in composition from the villages that were trained on.

A comparison of the means and standard deviations of radiance values in the four bands for positively identified village pixels and four savanna woodland pixels indicates that there is substantial overlap in the signatures of the two categories (Fig. 3). This phenomenon may be limited to the dry season, when apparently villages and naturally vegetated areas are characterized by similar proportions of bare soil and dry vegetation (grass, leaves, thatched roofs). The two categories may be more successfully separated in the wet season data. Reining and Egbert (1975), in similar work using a General Electric Image 100 system to identify villages in Niger, did not appear to encounter the same problem. It is suggested that the Niger study area, located in the Sahel savanna region, has sparser natural vegetation, that the natural vegetation-soil complex is of a type dissimilar to the village composition, or that villages there have a make-up significantly different from those studies here.

Despite the problems of separating villages from naturally vegetated areas, the classification results suggest the possibility of mapping and estimating the areal extent of natural vegetation and cultivated land by automated processing of LANDSAT data.

## CONCLUSIONS

This study, in addition to that of Reining and Egbert (1975), has demonstrated that the digital processing of LANDSAT data has potential for vegetation and settlement mapping in the West African savanna regions. At present the differentiation of villages from savanna woodland can be accomplished only by interpretation using spatial information

on size, shape, and location. Although such procedures are routine in air photograph interpretation and even in manual LANDSAT image interpretation, they are procedures we wish to avoid in automated processing largely because of difficulties of developing suitable algorithms for spatial pattern recognition. We believe it is probable that full classification, including resolution of the village/woodland confusion can be accomplished by use of a multitemporal analysis employing the information which would be available from the changes in vegetation through the seasons. In this study there were no good quality LANDSAT data available for the wet season, which precluded the use of the multitemporal method. The continued operation of LANDSAT I and II coupled with the anticipated increased data collection from the Kinshasa receiving station should, however, overcome this problem. The lack of up-to-date ground truth for this study area has hindered a thorough testing of the classification procedures. The analysis of remote sensing data must therefore be supplemented by appropriate field survey and more exact classification procedures for land characteristics.

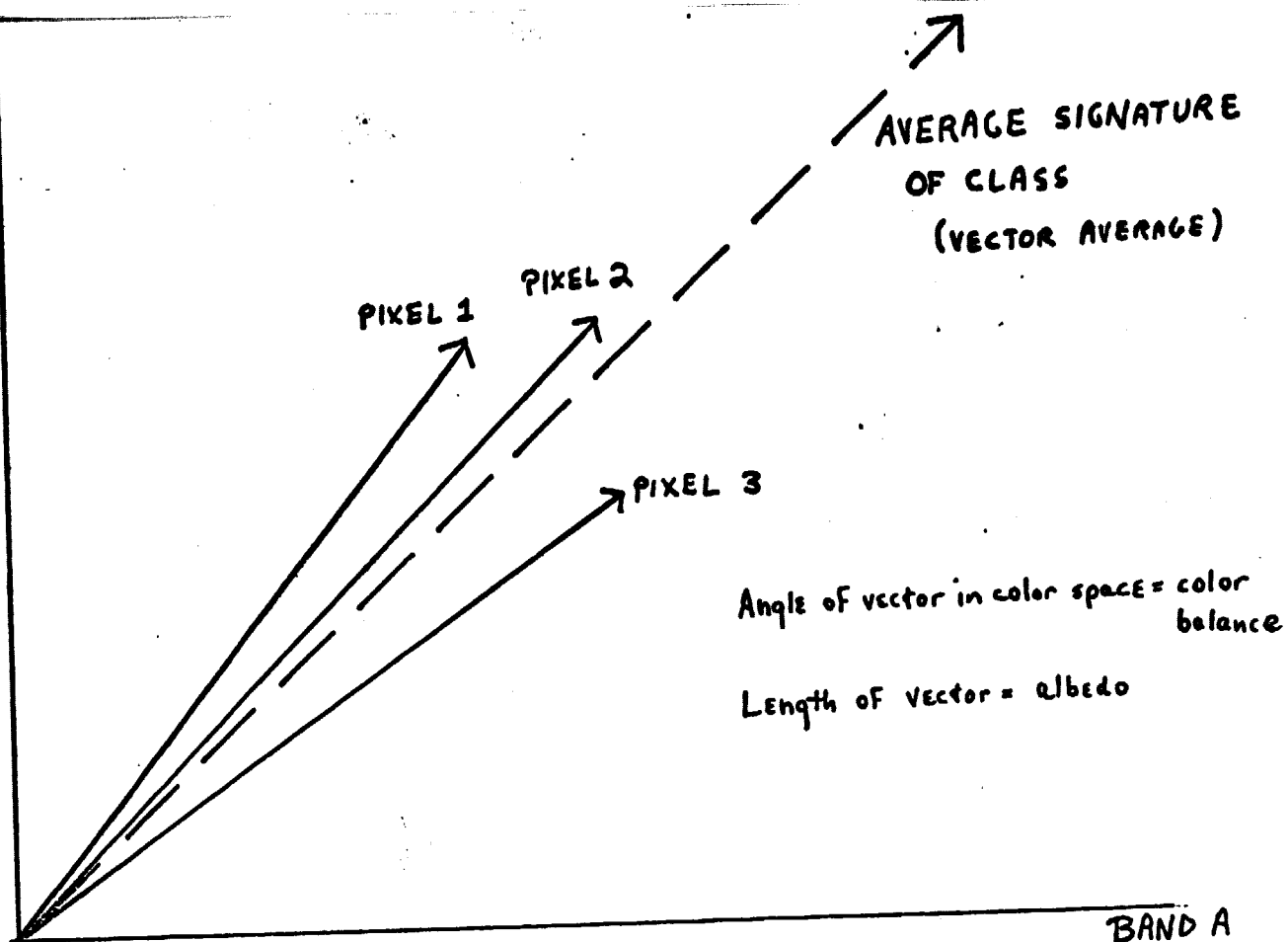
#### ACKNOWLEDGEMENT

This research was conducted under NASA grant No. NSG 5080, using facilities provided by the NASA Goddard Institute for Space Studies.

#### REFERENCE CITED

Reining, P. and D. Egbert. "Interactive multi-spectral analysis of more than one Sonrai village in Niger, West Africa." Paper presented at the NASA Earth Resources Survey Symposium, Houston, Texas, June, 1975.

BAND B



BAND B

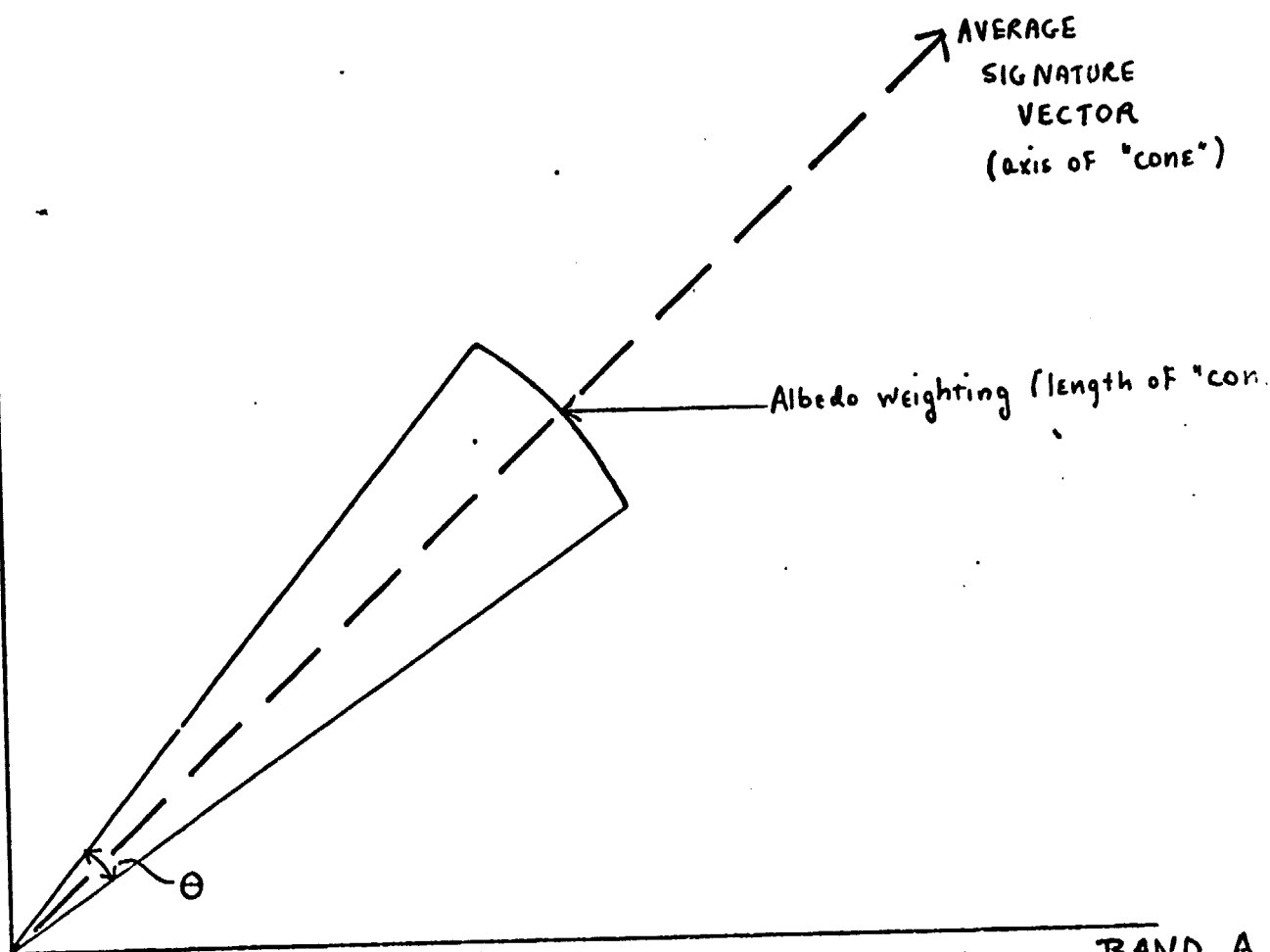


Fig. 1.

30

BAND A

TWO-COLOR SPACE REPRESENTATION OF GISS CLASSIFICATION ALGORITHM



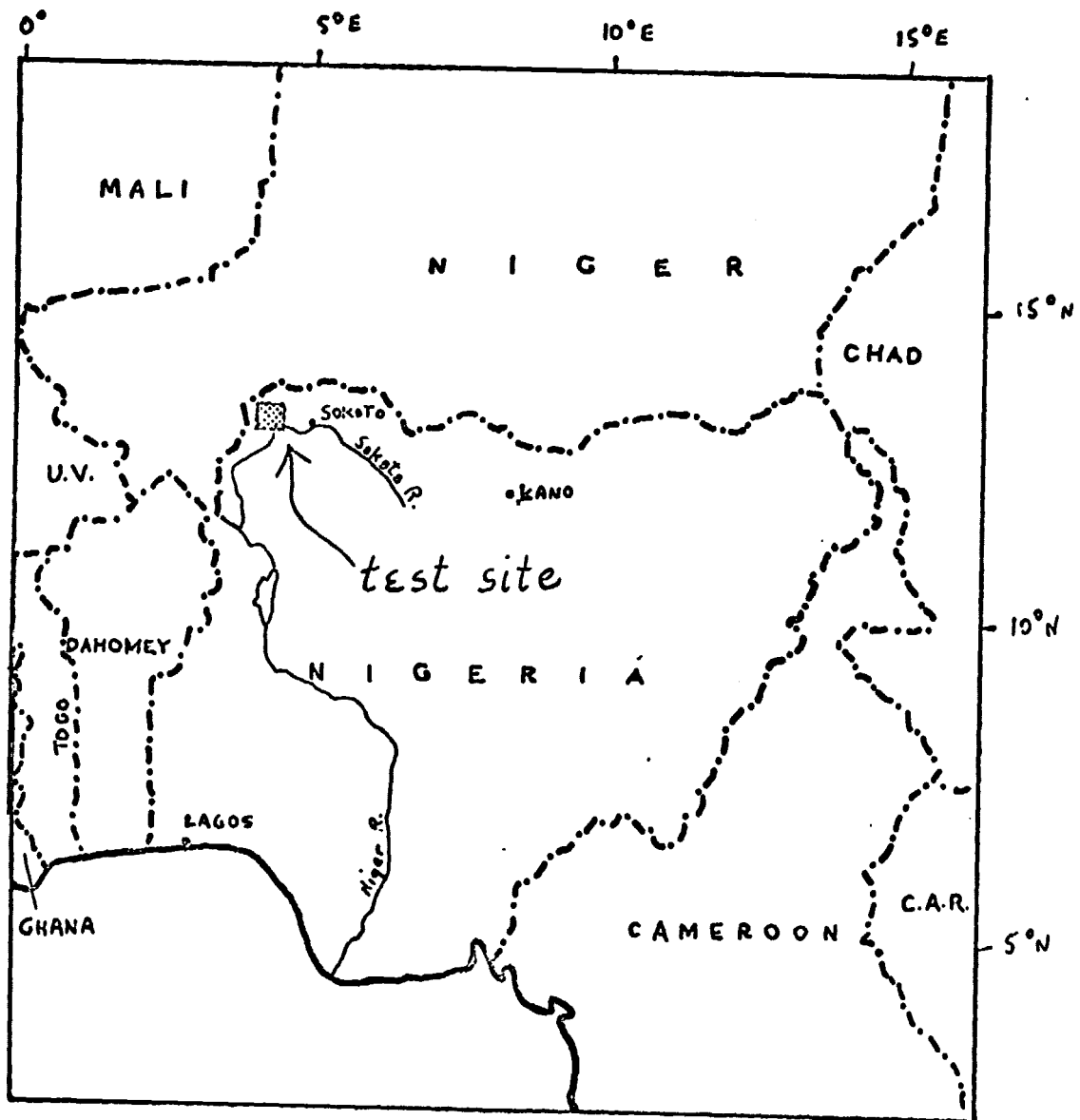


Fig. 2. Location of test site.

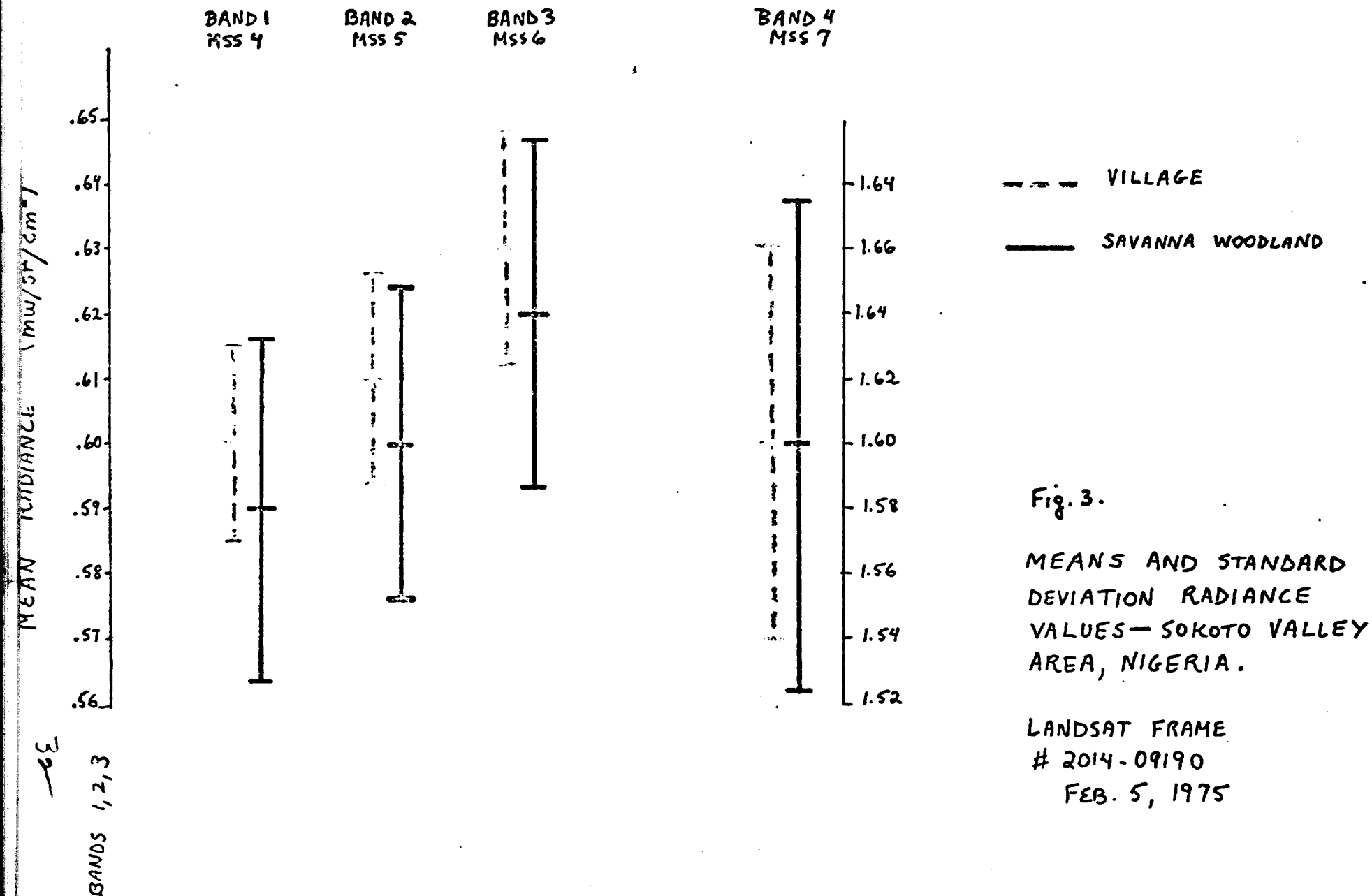


Fig. 4

SUMMARY OF POTENTIAL SUITABILITY OF LANDSAT DATA FOR  
IDENTIFICATION OF LAND COVER CLASSES IN DRY SAVANNA AREAS IN  
WEST AFRICA

	Identification Using Spectral Criteria	Identification Using Spatial Criteria
<u>LAND COVER CLASSES</u>		
<u>Settlement</u>		
1. Urban	good	not necessary
2. Villages	possible with multi-temporal data	necessary with uni-temporal dry season data
<u>Vegetation</u>		
1. Woodland Savanna	good (confusion with villages in uni-temporal dry season data)	not necessary if villages can be separately identified
2. Crop Land	good (potential confusion with barren land in dry season data)	not necessary if crop land can be separated from barren land
3. Fadama*	good	not necessary
<u>Other</u>		
1. Water	excellent	not necessary
2. Barren Land	good (may be some confusion with cleared crop land in uni-temporal dry season data)	not necessary if barren land can be separated from crop land

\*moist areas along streams or rivers which may be cropped nearly year-round